

AP20 Rec'd PCT/PTO 20 MAR 2006

- 1 -

SEQUENCE LISTING

<110> Bogosian, Gregg
 O'Neill, Julia P.
 Smith, Hong Q.

<120> Prevention of Incorporation of Non-Standard Amino Acids into Protein

<130> 11916.0059.00PC00

<150> US 60/505,807
 <151> 2003-09-25

<160> 16

<170> PatentIn version 3.3

<210> 1
 <211> 1344
 <212> DNA
 <213> Escherichia coli

<400> 1
 atggatcaga catattctct ggagtcattc ctcaaccatg tccaaaagcg cgacccgaat 60
 caaaccgagt tcgcgcaagc cgttcgtgaa gtaatgacca cactctggcc ttttcttgaa 120
 caaaatccaa aatatcgcca gatgtcatta ctggagcgctc tggttgaacc ggagcgcgctg 180
 atccagtttc gcgtgggtatg ggttgatgat cgcaaccaga tacagggtcaa ccgtgcatgg 240
 cgtgtgcagt tcagctctgc catcggcccg taaaaaggcg gtatgcgctt ccatccgtca 300
 gttaaccttt ccattctcaa attcctcggc tttgaacaaa ccttcaaaaa tgccctgact 360
 actctgccga tgggcgggtgg taaaggcggc agcgatttcg atccgaaagg aaaaagcgaa 420
 ggtgaagtga tgcgtttttg ccaggcgctg atgactgaac tgtatcgcca cctgggcgcg 480
 gataccgacg ttccggcagg tgatatcggg gttggtggtc gtgaagtcgg ctttatggcg 540
 gggatgatga aaaagctctc caacaatacc gcctgcgtct tcaccggtaa gggcctttca 600
 tttggcggca gtcttattcg cccggaagct accggctacg gtctggttta tttcacagaa 660
 gcaatgctaa aacgccacgg tatgggtttt gaagggatgc gcgtttccgt ttctggctcc 720
 ggcaacgtcg ccagtagcgc tategaaaaa gcgatggaat ttggtgctcg tgtgatcact 780
 gcgtcagact ccagcggcac tgtagttgat gaaagcggat tcacgaaaga gaaactggca 840
 cgtcttatcg aatcaaagc cagccgcgat ggtcgagtgg cagattacgc caaagaattt 900
 ggtctggtct atctcgaagg ccaacagccg tggtctctac cggttgatat cgccctgcct 960

- 2 -

```

tgcgccaccc agaatgaact ggatgttgac gccgcgcac acgttatcgc taatggcggt 1020
aaagccgtcg ccgaaggggc aaatatgccg accaccatcg aagcgactga actgttccag 1080
caggcaggcg tactatttgc accgggtaaa gcggctaata ctggtggcgt cgctacatcg 1140
ggcctggaaa tggcacaaaa cgctgcgcgc ctgggctgga aagccgagaa agttgacgca 1200
cgtttgcac acatcatgct ggatatccac catgcctgtg ttgagcatgg tgggaaggt 1260
gagcaaacca actacgtgca gggcgcgaa attgccgggt ttgtgaaggt tgccgatgcg 1320
atgctggcgc aggtgtgat ttaa 1344

```

```

<210> 2
<211> 447
<212> PRT
<213> Escherichia coli

```

```

<400> 2

```

```

Met Asp Gln Thr Tyr Ser Leu Glu Ser Phe Leu Asn His Val Gln Lys
1          5          10          15
Arg Asp Pro Asn Gln Thr Glu Phe Ala Gln Ala Val Arg Glu Val Met
          20          25          30
Thr Thr Leu Trp Pro Phe Leu Glu Gln Asn Pro Lys Tyr Arg Gln Met
          35          40          45
Ser Leu Leu Glu Arg Leu Val Glu Pro Glu Arg Val Ile Gln Phe Arg
          50          55          60
Val Val Trp Val Asp Asp Arg Asn Gln Ile Gln Val Asn Arg Ala Trp
65          70          75          80
Arg Val Gln Phe Ser Ser Ala Ile Gly Pro Tyr Lys Gly Gly Met Arg
          85          90          95
Phe His Pro Ser Val Asn Leu Ser Ile Leu Lys Phe Leu Gly Phe Glu
          100          105          110
Gln Thr Phe Lys Asn Ala Leu Thr Thr Leu Pro Met Gly Gly Gly Lys
          115          120          125
Gly Gly Ser Asp Phe Asp Pro Lys Gly Lys Ser Glu Gly Glu Val Met
          130          135          140
Arg Phe Cys Gln Ala Leu Met Thr Glu Leu Tyr Arg His Leu Gly Ala
145          150          155          160
Asp Thr Asp Val Pro Ala Gly Asp Ile Gly Val Gly Gly Arg Glu Val
          165          170          175

```

- 3 -

Gly Phe Met Ala Gly Met Met Lys Lys Leu Ser Asn Asn Thr Ala Cys
 180 185 190
 Val Phe Thr Gly Lys Gly Leu Ser Phe Gly Gly Ser Leu Ile Arg Pro
 195 200 205
 Glu Ala Thr Gly Tyr Gly Leu Val Tyr Phe Thr Glu Ala Met Leu Lys
 210 215 220
 Arg His Gly Met Gly Phe Glu Gly Met Arg Val Ser Val Ser Gly Ser
 225 230 235 240
 Gly Asn Val Ala Gln Tyr Ala Ile Glu Lys Ala Met Glu Phe Gly Ala
 245 250 255
 Arg Val Ile Thr Ala Ser Asp Ser Ser Gly Thr Val Val Asp Glu Ser
 260 265 270
 Gly Phe Thr Lys Glu Lys Leu Ala Arg Leu Ile Glu Ile Lys Ala Ser
 275 280 285
 Arg Asp Gly Arg Val Ala Asp Tyr Ala Lys Glu Phe Gly Leu Val Tyr
 290 295 300
 Leu Glu Gly Gln Gln Pro Trp Ser Leu Pro Val Asp Ile Ala Leu Pro
 305 310 315 320
 Cys Ala Thr Gln Asn Glu Leu Asp Val Asp Ala Ala His Gln Leu Ile
 325 330 335
 Ala Asn Gly Val Lys Ala Val Ala Glu Gly Ala Asn Met Pro Thr Thr
 340 345 350
 Ile Glu Ala Thr Glu Leu Phe Gln Gln Ala Gly Val Leu Phe Ala Pro
 355 360 365
 Gly Lys Ala Ala Asn Ala Gly Gly Val Ala Thr Ser Gly Leu Glu Met
 370 375 380
 Ala Gln Asn Ala Ala Arg Leu Gly Trp Lys Ala Glu Lys Val Asp Ala
 385 390 395 400
 Arg Leu His His Ile Met Leu Asp Ile His His Ala Cys Val Glu His
 405 410 415
 Gly Gly Glu Gly Glu Gln Thr Asn Tyr Val Gln Gly Ala Asn Ile Ala
 420 425 430
 Gly Phe Val Lys Val Ala Asp Ala Met Leu Ala Gln Gly Val Ile
 435 440 445

<210> 3

<211> 1344

<212> DNA

<213> Escherichia coli

- 4 -

<400> 3
 atggatcaga catattctct ggagtcattc ctcaaccatg tccaaaagcg cgacccgaat 60
 caaaccgagt tcgcgcaagc cgttcgtgaa gtaatgacca cactctggcc ttttcttgaa 120
 caaaatccaa aatatcgcca gatgtcatta ctggagcgtc tgggtgaacc ggagcgcggtg 180
 atccagtttc gcgtgggtatg ggttgatgat cgcaaccaga tacagggtcaa ccgtgcatgg 240
 cgtgtgcagt tcagctctgc catcggtccg tacctgggcg gtatgcgctt ccatccgtca 300
 gttaaccttt ccattctcaa attcctcggc tttgaacaaa ctttcaaaaa tgccctgact 360
 actctgccga tgggcgggtgg taaaggcggc agcgatttcg atccgaaagg aaaaagcgaa 420
 ggtgaagtga tgcgtttttg ccaggcgctg atgactgaac tgtatcgcca cctgggcgcg 480
 gataccgacg ttccggcagg tgatatcggg gttgggtggc gtgaagtcgg ctttatggcg 540
 gggatgatga aaaagctctc caacaatacc gcctgcgtct tcaccggtaa gggcctttca 600
 tttggcggca gtcttattcg cccggaagct accggctacg gtctggttta tttcacagaa 660
 gcaatgctaa aacgccacgg tatgggtttt gaagggatgc gcgtttccgt ttctggctcc 720
 ggcaacgtcg cccagtcacg tatcgaaaaa gcgatggaat ttggtgctcg tgtgatcact 780
 gcgtcagact ccagcggcac tgtagttgat gaaagcggat tcacgaaaga gaaactggca 840
 cgtcttatcg aaatcaaagc cagccgcgat ggtcgagtgg cagattacgc caaagaattt 900
 ggtctggtct atctcgaagg ccaacagccg tgggtctctac cggttgatat cgccctgcct 960
 tgcgccaccc agaatgaact ggatgttgac gccgcgcac agcttatcgc taatggcggt 1020
 aaagccgtcg ccgaaggggc aaatatgccg accaccatcg aagcgactga actgttccag 1080
 caggcaggcg tactatttgc accgggtaaa gcggctaag ctgggtggcg cgctacatcg 1140
 ggcctggaaa tggcacaaaa cgctgcgcgc ctgggctgga aagccgagaa agttgacgca 1200
 cgtttgcac acatcatgct ggatatccac catgcctgtg ttgagcatgg tgggtgaagg 1260
 gagcaaacca actacgtgca gggcgcgaa attgccggtt ttgtgaagg tgccgatgcg 1320
 atgctggcgc aggggtgtgat ttaa 1344

<210> 4
 <211> 447
 <212> PRT
 <213> Escherichia coli

<400> 4

- 5 -

Met Asp Gln Thr Tyr Ser Leu Glu Ser Phe Leu Asn His Val Gln Lys
 1 5 10 15
 Arg Asp Pro Asn Gln Thr Glu Phe Ala Gln Ala Val Arg Glu Val Met
 20 25 30
 Thr Thr Leu Trp Pro Phe Leu Glu Gln Asn Pro Lys Tyr Arg Gln Met
 35 40 45
 Ser Leu Leu Glu Arg Leu Val Glu Pro Glu Arg Val Ile Gln Phe Arg
 50 55 60
 Val Val Trp Val Asp Asp Arg Asn Gln Ile Gln Val Asn Arg Ala Trp
 65 70 75 80
 Arg Val Gln Phe Ser Ser Ala Ile Gly Pro Tyr Leu Gly Gly Met Arg
 85 90 95
 Phe His Pro Ser Val Asn Leu Ser Ile Leu Lys Phe Leu Gly Phe Glu
 100 105 110
 Gln Thr Phe Lys Asn Ala Leu Thr Thr Leu Pro Met Gly Gly Gly Lys
 115 120 125
 Gly Gly Ser Asp Phe Asp Pro Lys Gly Lys Ser Glu Gly Glu Val Met
 130 135 140
 Arg Phe Cys Gln Ala Leu Met Thr Glu Leu Tyr Arg His Leu Gly Ala
 145 150 155 160
 Asp Thr Asp Val Pro Ala Gly Asp Ile Gly Val Gly Gly Arg Glu Val
 165 170 175
 Gly Phe Met Ala Gly Met Met Lys Lys Leu Ser Asn Asn Thr Ala Cys
 180 185 190
 Val Phe Thr Gly Lys Gly Leu Ser Phe Gly Gly Ser Leu Ile Arg Pro
 195 200 205
 Glu Ala Thr Gly Tyr Gly Leu Val Tyr Phe Thr Glu Ala Met Leu Lys
 210 215 220
 Arg His Gly Met Gly Phe Glu Gly Met Arg Val Ser Val Ser Gly Ser
 225 230 235 240
 Gly Asn Val Ala Gln Tyr Ala Ile Glu Lys Ala Met Glu Phe Gly Ala
 245 250 255
 Arg Val Ile Thr Ala Ser Asp Ser Ser Gly Thr Val Val Asp Glu Ser
 260 265 270
 Gly Phe Thr Lys Glu Lys Leu Ala Arg Leu Ile Glu Ile Lys Ala Ser
 275 280 285
 Arg Asp Gly Arg Val Ala Asp Tyr Ala Lys Glu Phe Gly Leu Val Tyr

- 6 -

290	295	300
Leu Glu Gly Gln Gln Pro Trp Ser Leu Pro Val Asp Ile Ala Leu Pro		
305	310	315 320
Cys Ala Thr Gln Asn Glu Leu Asp Val Asp Ala Ala His Gln Leu Ile		
	325	330 335
Ala Asn Gly Val Lys Ala Val Ala Glu Gly Ala Asn Met Pro Thr Thr		
	340	345 350
Ile Glu Ala Thr Glu Leu Phe Gln Gln Ala Gly Val Leu Phe Ala Pro		
	355	360 365
Gly Lys Ala Ala Asn Ala Gly Gly Val Ala Thr Ser Gly Leu Glu Met		
	370	375 380
Ala Gln Asn Ala Ala Arg Leu Gly Trp Lys Ala Glu Lys Val Asp Ala		
	385	390 395 400
Arg Leu His His Ile Met Leu Asp Ile His His Ala Cys Val Glu His		
	405	410 415
Gly Gly Glu Gly Glu Gln Thr Asn Tyr Val Gln Gly Ala Asn Ile Ala		
	420	425 430
Gly Phe Val Lys Val Ala Asp Ala Met Leu Ala Gln Gly Val Ile		
	435	440 445

<210> 5
 <211> 1101
 <212> DNA
 <213> Bacillus cereus

<400> 5
 atgacattag aaatcttcga atacttagaa aaatatgatt atgagcaagt agtattttgt 60
 caagataaag aatctggttt aaaagcaatt attgcaattc atgatacaac acttggaaccg 120
 gctcttggtg gaacaagaat gtggacatat gattctgaag aagcggcgat tgaagatgca 180
 ttgcgtcttg caaaagggat gacatacaaa aacgcagcag ctggttttaa cttaggtggt 240
 gcgaaaacag taattatcgg tgatcctcgt aaagataaga gcgaagcaat gttccgtgca 300
 ctaggacggt atatccaagg actaaacgga cgttacatta cagctgaaga tgttggtaca 360
 acagtagatg atatggatat tatccatgaa gaaactgact ttgtaacagg tatctcacca 420
 tcattcgggt cttctggtaa cccatctccg gtaactgcat acggtgttta ccgtgggatg 480
 aaagcagctg caaaagaagc tttcgggtact gacaatttag aaggaaaagt aattgctggt 540
 caaggcggtg gtaacgtagc atatcaccta tgcaaacatt tacacgctga aggagcaaaa 600

- 7 -

```

ttaatcggtta cagatattaa taaagaagct gtacaacgtg ctgtagaaga attcgggtgca      660
tcagcagttg aaccaaataa aatttacggt gttgaatgcg atatttacgc accatgtgca      720
ctaggcgcaa cagttaatga tgaaactatt ccacaactta aagcaaaagt aatcgcaggt      780
tctgcaaata accaattaaa agaagatcgt catggtgaca tcattcatga aatgggtatt      840
gtatacgcac cagattatgt aattaatgca ggtggcgtaa ttaacgtagc agacgaatta      900
tatggatata atagagaacg tgcactaaaa cgtggttgagt ctatttatga cagcattgca      960
aaagtaatcg aaatttcaaa acgcgatggc atagcaactt atgtagcggc agatcgtcta     1020
gctgaagagc gcattgcaag cttgaagaat tctcgtagca cttacttacg caacgggtcac     1080
gatattatta gccgtcgcta a                                             1101

```

```

<210> 6
<211> 366
<212> PRT
<213> Bacillus cereus

```

```

<400> 6

```

```

Met Thr Leu Glu Ile Phe Glu Tyr Leu Glu Lys Tyr Asp Tyr Glu Gln
1           5           10           15
Val Val Phe Cys Gln Asp Lys Glu Ser Gly Leu Lys Ala Ile Ile Ala
20           25           30
Ile His Asp Thr Thr Leu Gly Pro Ala Leu Gly Gly Thr Arg Met Trp
35           40           45
Thr Tyr Asp Ser Glu Glu Ala Ala Ile Glu Asp Ala Leu Arg Leu Ala
50           55           60
Lys Gly Met Thr Tyr Lys Asn Ala Ala Ala Gly Leu Asn Leu Gly Gly
65           70           75           80
Ala Lys Thr Val Ile Ile Gly Asp Pro Arg Lys Asp Lys Ser Glu Ala
85           90           95
Met Phe Arg Ala Leu Gly Arg Tyr Ile Gln Gly Leu Asn Gly Arg Tyr
100          105          110
Ile Thr Ala Glu Asp Val Gly Thr Thr Val Asp Asp Met Asp Ile Ile
115          120          125
His Glu Glu Thr Asp Phe Val Thr Gly Ile Ser Pro Ser Phe Gly Ser
130          135          140
Ser Gly Asn Pro Ser Pro Val Thr Ala Tyr Gly Val Tyr Arg Gly Met
145          150          155          160

```

- 8 -

Lys Ala Ala Ala Lys Glu Ala Phe Gly Thr Asp Asn Leu Glu Gly Lys
 165 170 175
 Val Ile Ala Val Gln Gly Val Gly Asn Val Ala Tyr His Leu Cys Lys
 180 185 190
 His Leu His Ala Glu Gly Ala Lys Leu Ile Val Thr Asp Ile Asn Lys
 195 200 205
 Glu Ala Val Gln Arg Ala Val Glu Glu Phe Gly Ala Ser Ala Val Glu
 210 215 220
 Pro Asn Glu Ile Tyr Gly Val Glu Cys Asp Ile Tyr Ala Pro Cys Ala
 225 230 235 240
 Leu Gly Ala Thr Val Asn Asp Glu Thr Ile Pro Gln Leu Lys Ala Lys
 245 250 255
 Val Ile Ala Gly Ser Ala Asn Asn Gln Leu Lys Glu Asp Arg His Gly
 260 265 270
 Asp Ile Ile His Glu Met Gly Ile Val Tyr Ala Pro Asp Tyr Val Ile
 275 280 285
 Asn Ala Gly Gly Val Ile Asn Val Ala Asp Glu Leu Tyr Gly Tyr Asn
 290 295 300
 Arg Glu Arg Ala Leu Lys Arg Val Glu Ser Ile Tyr Asp Thr Ile Ala
 305 310 315 320
 Lys Val Ile Glu Ile Ser Lys Arg Asp Gly Ile Ala Thr Tyr Val Ala
 325 330 335
 Ala Asp Arg Leu Ala Glu Glu Arg Ile Ala Ser Leu Lys Asn Ser Arg
 340 345 350
 Ser Thr Tyr Leu Arg Asn Gly His Asp Ile Ile Ser Arg Arg
 355 360 365
 <210> 7
 <211> 1098
 <212> DNA
 <213> *Bacillus subtilis*
 <400> 7
 atggaacttt ttaaataatat ggagaaatac gattatgaac aattagtatt ctgccaagat 60
 gaacaatcag gcttaaaagc gattatcgca attcatgata caacgcttgg tccggcgctt 120
 ggcggaacga gaatgtggac atatgaaaat gaagaagcgg caattgaaga cgcgctcaga 180
 ctggcaagag gcatgaccta taaagacgcg gctgcaggcc taaaccttgg cggcggaaaa 240
 acagtaataa tcggcgatcc acgcaaagac aaaaatgaag aaatgttccg cgcgtttggc 300

- 9 -

cgctatatattc aaggactgaa cggcagatac attacagctg aagatgtggg tacaacgggt 360
 gaggatatgg acattattca tgatgaaaca gactatgtca cagggatttc tcctgctttc 420
 ggctcttctg gaaatccatc tccagttaca gcgtacgggg tgtacagagg aatgaaagca 480
 gccgctaaag ctgctttcgg aaccgactct cttgaaggga aaaccatcgc tgtacagggt 540
 gtaggggaatg tggcctacaa cctatgccgc cacctgcatg aagaaggggc aaacttaatc 600
 gttacgggata tcaacaaaca gtcagtacag cgcgcagttg aagatttttg cgcctgtgcg 660
 gttgatccgg aagagattta ttcacaagag tgcgatattt atgctccgtg cgcccttgga 720
 gcgacaatca acgacgacac cattaacacag ctgaaggcga aagtcacgc ggggtgcggct 780
 aataaccaat taaaagaaac gcgccatggc gatcaaattc acgaaatggg tatcgtttac 840
 gcaccgggatt atgtcattaa cgcaggcggc gtgatcaacg tggcagatga gctttacggc 900
 tataatgcag aacgtgcatt gaaaaaagtt gaaggcattt acggcaatat tgagcgtgta 960
 cttgagattt ctcagcgtga cggcattcca acatatttag cagctgaccg cttggcagag 1020
 gaacggattg aacgcacgac ccgctcaaga agccagtttt tgcaaaacgg ccacagtgtgta 1080
 ttaagcagac gttaatatag 1098

<210> 8
 <211> 364
 <212> PRT
 <213> Bacillus subtilis

<400> 8

Met Glu Leu Phe Lys Tyr Met Glu Lys Tyr Asp Tyr Glu Gln Leu Val
 1 5 10 15
 Phe Cys Gln Asp Glu Gln Ser Gly Leu Lys Ala Ile Ile Ala Ile His
 20 25 30
 Asp Thr Thr Leu Gly Pro Ala Leu Gly Gly Thr Arg Met Trp Thr Tyr
 35 40 45
 Glu Asn Glu Glu Ala Ala Ile Glu Asp Ala Leu Arg Leu Ala Arg Gly
 50 55 60
 Met Thr Tyr Lys Asp Ala Ala Ala Gly Leu Asn Leu Gly Gly Gly Lys
 65 70 75 80
 Thr Val Ile Ile Gly Asp Pro Arg Lys Asp Lys Asn Glu Glu Met Phe
 85 90 95
 Arg Ala Phe Gly Arg Tyr Ile Gln Gly Leu Asn Gly Arg Tyr Ile Thr
 100 105 110

- 10 -

Ala Glu Asp Val Gly Thr Thr Val Glu Asp Met Asp Ile Ile His Asp
 115 120 125
 Glu Thr Asp Tyr Val Thr Gly Ile Ser Pro Ala Phe Gly Ser Ser Gly
 130 135 140
 Asn Pro Ser Pro Val Thr Ala Tyr Gly Val Tyr Arg Gly Met Lys Ala
 145 150 155 160
 Ala Ala Lys Ala Ala Phe Gly Thr Asp Ser Leu Glu Gly Lys Thr Ile
 165 170 175
 Ala Val Gln Gly Val Gly Asn Val Ala Tyr Asn Leu Cys Arg His Leu
 180 185 190
 His Glu Glu Gly Ala Asn Leu Ile Val Thr Asp Ile Asn Lys Gln Ser
 195 200 205
 Val Gln Arg Ala Val Glu Asp Phe Gly Ala Arg Ala Val Asp Pro Glu
 210 215 220
 Glu Ile Tyr Ser Gln Glu Cys Asp Ile Tyr Ala Pro Cys Ala Leu Gly
 225 230 235 240
 Ala Thr Ile Asn Asp Asp Thr Ile Lys Gln Leu Lys Ala Lys Val Ile
 245 250 255
 Ala Gly Ala Ala Asn Asn Gln Leu Lys Glu Thr Arg His Gly Asp Gln
 260 265 270
 Ile His Glu Met Gly Ile Val Tyr Ala Pro Asp Tyr Val Ile Asn Ala
 275 280 285
 Gly Gly Val Ile Asn Val Ala Asp Glu Leu Tyr Gly Tyr Asn Ala Glu
 290 295 300
 Arg Ala Leu Lys Lys Val Glu Gly Ile Tyr Gly Asn Ile Glu Arg Val
 305 310 315 320
 Leu Glu Ile Ser Gln Arg Asp Gly Ile Pro Thr Tyr Leu Ala Ala Asp
 325 330 335
 Arg Leu Ala Glu Glu Arg Ile Glu Arg Met Arg Arg Ser Arg Ser Gln
 340 345 350
 Phe Leu Gln Asn Gly His Ser Val Leu Ser Arg Arg
 355 360

<210> 9
 <211> 1062
 <212> DNA
 <213> Nostoc sp.

<400> 9

- 11 -

```

atgcagctat ttgaaactgt tagagaaatg ggacatgagc aagtactcta ctgtcatgga      60
aaaaatccag atattagagc aataattgcc atccatgaca ccacattagg cccagcaatg      120
ggagccacaa ggctttatcc ttatatcaat gaagaagccg ccttaagaga tgctttgcgt      180
ttgagtcggg ggatgactta taaagcagct tgcgctaaca ttcccgcagg cggaggcaaa      240
gccgttatta ttgccaatcc cgaagataaa acagatgaaa tgttgagagc ttatggacgc      300
tttgtggaaa gtctcaaagg tagatttatt accgggcaag atgtgaatat cactccacaa      360
gatgtccgca caattaaaca agaaaccaat tatgtagttg gtgtggaaga aaaatctggt      420
gggcctgctc ctatcacagc tttaggcgta tttttaggta ttaaagctgc tgtagaattt      480
cgctggcaaa ctaaaaatat tgaagggatg acagttgccg ttcaaggttt aggaaatggt      540
ggtcagaatc tctgccgaca cttacatgaa aatgggtataa agcttatagt tgctgatttt      600
agttctgaaa aaacagcaga aataaaacac ctttttggtg ctacagtagt agagccagat      660
gaaattttact cacaaaatgt agacatatTT tctccctgtg ctatgggagg aattattaac      720
agtcaaacaa ttcccact acaagccaaa attattgctg gtgctgccaa taaccagtta      780
gataatgagc gtctgcatgg tcaaagatta gtagaaaaag atatcctcta ctgtcctgat      840
tatgtaatca atgctggtgg tatcatcaac gtttataacg aaatgattgg ctatgaagaa      900
gataaggcct tcaagcaagt taataatatt tacgacacat tattagcaat tttcaatatt      960
gctcaacaac aaagcattac tactaatgat gcttcaaaac ggcttgacaga tgaaaggatt     1020
atgaaggcga gaatcaataa aaatcaacta attgctgcct aa                        1062

```

<210> 10
 <211> 353
 <212> PRT
 <213> Nostoc sp.

<400> 10

```

Met Gln Leu Phe Glu Thr Val Arg Glu Met Gly His Glu Gln Val Leu
1              5              10              15
Tyr Cys His Gly Lys Asn Pro Asp Ile Arg Ala Ile Ile Ala Ile His
                20              25              30
Asp Thr Thr Leu Gly Pro Ala Met Gly Ala Thr Arg Leu Tyr Pro Tyr
              35              40              45
Ile Asn Glu Glu Ala Ala Leu Arg Asp Ala Leu Arg Leu Ser Arg Gly
50              55              60

```

- 12 -

Met Thr Tyr Lys Ala Ala Cys Ala Asn Ile Pro Ala Gly Gly Gly Lys
 65 70 75 80
 Ala Val Ile Ile Ala Asn Pro Glu Asp Lys Thr Asp Glu Met Leu Arg
 85 90 95
 Ala Tyr Gly Arg Phe Val Glu Ser Leu Lys Gly Arg Phe Ile Thr Gly
 100 105 110
 Gln Asp Val Asn Ile Thr Pro Gln Asp Val Arg Thr Ile Lys Gln Glu
 115 120 125
 Thr Asn Tyr Val Val Gly Val Glu Glu Lys Ser Gly Gly Pro Ala Pro
 130 135 140
 Ile Thr Ala Leu Gly Val Phe Leu Gly Ile Lys Ala Ala Val Glu Phe
 145 150 155 160
 Arg Trp Gln Thr Lys Asn Ile Glu Gly Met Thr Val Ala Val Gln Gly
 165 170 175
 Leu Gly Asn Val Gly Gln Asn Leu Cys Arg His Leu His Glu Asn Gly
 180 185 190
 Ile Lys Leu Ile Val Ala Asp Phe Ser Ser Glu Lys Thr Ala Glu Ile
 195 200 205
 Lys His Leu Phe Gly Ala Thr Val Val Glu Pro Asp Glu Ile Tyr Ser
 210 215 220
 Gln Asn Val Asp Ile Phe Ser Pro Cys Ala Met Gly Gly Ile Ile Asn
 225 230 235 240
 Ser Gln Thr Ile Pro Gln Leu Gln Ala Lys Ile Ile Ala Gly Ala Ala
 245 250 255
 Asn Asn Gln Leu Asp Asn Glu Arg Leu His Gly Gln Arg Leu Val Glu
 260 265 270
 Lys Asp Ile Leu Tyr Cys Pro Asp Tyr Val Ile Asn Ala Gly Gly Ile
 275 280 285
 Ile Asn Val Tyr Asn Glu Met Ile Gly Tyr Glu Glu Asp Lys Ala Phe
 290 295 300
 Lys Gln Val Asn Asn Ile Tyr Asp Thr Leu Leu Ala Ile Phe Asn Ile
 305 310 315 320
 Ala Gln Gln Gln Ser Ile Thr Thr Asn Asp Ala Ser Lys Arg Leu Ala
 325 330 335
 Asp Glu Arg Ile Met Lys Ala Arg Ile Asn Lys Asn Gln Leu Ile Ala
 340 345 350
 Ala

- 13 -

<210> 11
 <211> 1044
 <212> DNA
 <213> *Shewanella oneidensis*

<400> 11
 atggctgtat ttaatcatgt atcctttgat gagcatgaac aggtcgtatt ctgtcatgat 60
 aaagaaagtg gcttaaaagc cattattgcc atccataata ccaatttagg ccctgctgtg 120
 ggtggatgcc ggatgtggaa ctaccaatcc gatgacgaag ccctgacaga cgtattacgc 180
 ctctcccgtg gtatgactta caaaaacgcg ctcgctgggt taaccatggg cgggtggtaaa 240
 tcagtgatta ttgccgatcc taagcgcctt gaccgcgaag ccctcttccg tgcttttggc 300
 cgttttatca atagtctcgg tggacgttac tattccgcag aagacgttgg caccacgaca 360
 gctgatatta tgatcgccca tcaagaaacg ccctatatgg cggggccttg aggcaagagt 420
 ggcgatcctt ctccgtttac ggcactaggt acttatttag gtatcaaggc cgcggttaaa 480
 cataagctcg atttagacag cttaaagggc cttaagatcg ccgttcaagg tgttggccat 540
 gtgggttatt atctgtgtaa acatctacat gaagaagggt cacagctaatt tgttaccgat 600
 attcatcagg cgtcacttga taaagtggct accgactttg gtgctaccgt tgttgcacca 660
 caggatatct acgccaaga cgtcgatgtg tacgccccat gcgcactagg tgcgacctta 720
 aacgatgtta ccctgccact actcaaagct aagattgttg caggttgtgc caacaaccaa 780
 ttagccgaag tacgccatgg cgagcagtta aaagaaatgg gcattcttta tgcgccagat 840
 tatgtgatta acgcgggcgg cattattaac gtatcattcg aaaaagacta tgatgcggcg 900
 aaatcagaag ctaaggtcag agaaatctac aacacgctgc tgaagatttt tgctaaagcc 960
 gatgctgaga accgcacgac aggtgcagta gctgacgaaa tggcccgtgc aatttaccaa 1020
 gcgccaaagc ctaatagggc ttag 1044

<210> 12
 <211> 347
 <212> PRT
 <213> *Shewanella oneidensis*

<400> 12

Met Ala Val Phe Asn His Val Ser Phe Asp Glu His Glu Gln Val Val
 1 5 10 15
 Phe Cys His Asp Lys Glu Ser Gly Leu Lys Ala Ile Ile Ala Ile His

- 14 -

20					25					30					
Asn	Thr	Asn	Leu	Gly	Pro	Ala	Val	Gly	Gly	Cys	Arg	Met	Trp	Asn	Tyr
		35					40					45			
Gln	Ser	Asp	Asp	Glu	Ala	Leu	Thr	Asp	Val	Leu	Arg	Leu	Ser	Arg	Gly
	50					55					60				
Met	Thr	Tyr	Lys	Asn	Ala	Leu	Ala	Gly	Leu	Thr	Met	Gly	Gly	Gly	Lys
65						70					75				80
Ser	Val	Ile	Ile	Ala	Asp	Pro	Lys	Arg	Pro	Asp	Arg	Glu	Ala	Leu	Phe
				85					90					95	
Arg	Ala	Phe	Gly	Arg	Phe	Ile	Asn	Ser	Leu	Gly	Gly	Arg	Tyr	Tyr	Ser
			100					105					110		
Ala	Glu	Asp	Val	Gly	Thr	Thr	Thr	Ala	Asp	Ile	Met	Ile	Ala	His	Gln
		115					120					125			
Glu	Thr	Pro	Tyr	Met	Ala	Gly	Leu	Glu	Gly	Lys	Ser	Gly	Asp	Pro	Ser
	130					135					140				
Pro	Phe	Thr	Ala	Leu	Gly	Thr	Tyr	Leu	Gly	Ile	Lys	Ala	Ala	Val	Lys
145						150					155				160
His	Lys	Leu	Asp	Leu	Asp	Ser	Leu	Lys	Gly	Leu	Lys	Ile	Ala	Val	Gln
				165					170					175	
Gly	Val	Gly	His	Val	Gly	Tyr	Tyr	Leu	Cys	Lys	His	Leu	His	Glu	Glu
			180					185					190		
Gly	Ala	Gln	Leu	Ile	Val	Thr	Asp	Ile	His	Gln	Ala	Ser	Leu	Asp	Lys
		195					200					205			
Val	Ala	Thr	Asp	Phe	Gly	Ala	Thr	Val	Val	Ala	Pro	Gln	Asp	Ile	Tyr
	210					215					220				
Ala	Gln	Asp	Val	Asp	Val	Tyr	Ala	Pro	Cys	Ala	Leu	Gly	Ala	Thr	Leu
225						230					235				240
Asn	Asp	Val	Thr	Leu	Pro	Leu	Leu	Lys	Ala	Lys	Ile	Val	Ala	Gly	Cys
				245					250					255	
Ala	Asn	Asn	Gln	Leu	Ala	Glu	Val	Arg	His	Gly	Glu	Gln	Leu	Lys	Glu
			260					265					270		
Met	Gly	Ile	Leu	Tyr	Ala	Pro	Asp	Tyr	Val	Ile	Asn	Ala	Gly	Gly	Ile
	275						280					285			
Ile	Asn	Val	Ser	Phe	Glu	Lys	Asp	Tyr	Asp	Ala	Ala	Lys	Ser	Glu	Ala
	290						295				300				
Lys	Val	Arg	Glu	Ile	Tyr	Asn	Thr	Leu	Leu	Lys	Ile	Phe	Ala	Lys	Ala
305						310					315				320

```
<210> 13
<211> 1086
<212> DNA
<213> Streptomyces avermitilis
```

```
<210> 14
<211> 361
<212> PRT
<213> Streptomyces avermitilis
```

- 16 -

<400> 14

Met	Thr	Asp	Val	Ser	Asp	Gly	Val	Leu	His	Thr	Leu	Phe	Arg	Ser	Asp	1	5	10	15
Gln	Gly	Gly	His	Glu	Gln	Val	Val	Leu	Cys	Gln	Asp	Arg	Ala	Thr	Gly	20	25	30	
Leu	Lys	Ala	Val	Ile	Ala	Ile	His	Ser	Thr	Ala	Leu	Gly	Pro	Ala	Leu	35	40	45	
Gly	Gly	Thr	Arg	Phe	Tyr	Pro	Tyr	Ala	Ser	Glu	Glu	Glu	Ala	Val	Ala	50	55	60	
Asp	Ala	Leu	Asn	Leu	Ala	Arg	Gly	Met	Ser	Tyr	Lys	Asn	Ala	Met	Ala	65	70	75	80
Gly	Leu	Asp	His	Gly	Gly	Gly	Lys	Ala	Val	Ile	Ile	Gly	Asp	Pro	Glu	85	90	95	
Arg	Ile	Lys	Thr	Glu	Glu	Leu	Leu	Leu	Ala	Tyr	Gly	Arg	Phe	Val	Ala	100	105	110	
Ser	Leu	Gly	Gly	Arg	Tyr	Val	Thr	Ala	Cys	Asp	Val	Gly	Thr	Tyr	Val	115	120	125	
Ala	Asp	Met	Asp	Val	Val	Ala	Arg	Glu	Cys	Arg	Trp	Thr	Thr	Gly	Arg	130	135	140	
Ser	Pro	Glu	Asn	Gly	Gly	Ala	Gly	Asp	Ser	Ser	Val	Leu	Thr	Ala	Phe	145	150	155	160
Gly	Val	Phe	Gln	Gly	Met	Arg	Ala	Ser	Ala	Gln	His	Leu	Trp	Gly	Asp	165	170	175	
Pro	Thr	Leu	Arg	Gly	Arg	Lys	Val	Gly	Ile	Ala	Gly	Val	Gly	Lys	Val	180	185	190	
Gly	Arg	His	Leu	Val	Arg	His	Leu	Leu	Asp	Asp	Gly	Ala	Glu	Val	Val	195	200	205	
Ile	Thr	Asp	Val	Arg	Thr	Asp	Ser	Val	Gln	Arg	Ile	Leu	Asp	Gln	His	210	215	220	
Pro	Thr	Gly	Val	Thr	Ala	Val	Ala	Asp	Thr	Asp	Ala	Leu	Ile	Arg	Val	225	230	235	240
Asp	Gly	Leu	Asp	Ile	Tyr	Ala	Pro	Cys	Ala	Leu	Gly	Gly	Ala	Leu	Asn	245	250	255	
Asp	Asp	Ser	Val	Thr	Val	Leu	Thr	Ala	Lys	Ile	Val	Cys	Gly	Ala	Ala	260	265	270	
Asn	Asn	Gln	Leu	Ala	His	Thr	Gly	Val	Glu	Lys	Asp	Leu	Ala	Asp	Arg				

- 17 -

275	280	285	
Gly Ile Leu Tyr Ala Pro Asp Tyr Val Val Asn Ala Gly Gly Val Ile			
290	295	300	
Gln Val Ala Asp Glu Leu His Gly Phe Asp Phe Asp Arg Cys Lys Ala			
305	310	315	320
Lys Ala Ala Lys Ile Phe Asp Thr Thr Leu Ala Ile Phe Ala Arg Ala			
	325	330	335
Lys Glu Asp Gly Ile Pro Pro Ala Ala Ala Asp Arg Ile Ala Glu			
	340	345	350
Gln Arg Met Ala Glu Ala Arg Arg Gly			
	355	360	
<210> 15			
<211> 1347			
<212> DNA			
<213> Nitrosomonas europaea			
<400> 15			
atgaaataca acagtatcga ggaattcaag aattatgttt ccgaaaggaa tccagggcaa			60
cccgaattcc tgcaggccgt ttcagaagtc attgaaagct tgtggccttt tatcgtcgat			120
cattctcggtt acgctgagca ggggttgctg gatcggctga tcgagccgga gcgcatgata			180
atattccggg tggcgtgggt ggatgatcgg ggtgaagtca aggtcaatcg ggggtaccgc			240
attcaatata attcggcgat cggccccatac aaggggggta cgcgcttcca tccgtcagtc			300
aacctttcca ttctcaaatt ccttgcattt gagcagactt tcaagaatgc actgacaaca			360
ttgccgatgg gaggaggcaa ggggtgatcg gattttgatc ccaagggtaa aagtcccggg			420
gaaatcatgc gcttctgcc aagcgtatgcg gccgaactgt tccggcatgt cgggtcggat			480
acggatgtac ctgccggaga catcgggtgtg ggcggacggg aagtcggcta catggctggt			540
atggtcaaga agctgaccaa ccgttcggac tgtgtattta ccggcaaagg attgaccttc			600
gggggatcgc tgctgcggcc ggaagctacc gggtagggtc tggcttattt tgccgaagag			660
atgctgaatc actccgggtt ttcatgaaa ggcatgcggg tatccgtatc cggttccggg			720
aacgtggcac agtttgccat tgacaaggcc atgtcgctgg gtgccaaagt agtcacgggt			780
tcagattcga gtgggtacggg ggtggatgaa gccggtttta caccagaaaa actggcaatt			840
ctggccgaag tcaagaatcg tctctacggg cgtgtcaatg aatttgctga acgggtggaa			900
gcacagttcc ttccgggtga aaaaccgtgg catgtgccgg tggatgtcgc tttgcctgt			960

- 18 -

gcgacccaga atgaactgaa cgaaaacgac gccgcaatac tgatcaggaa tgggtgcgaat 1020
 tgtgtggccg aggggtgcaa tatgccatgc actgcaggtg ccgtggaacg attccatcat 1080
 gcgaaagtac tgtttgcacc tggcaaggcg agcaacgcag gcggagtggc tacctcgggt 1140
 ctggaaatga gccagcaggc catgcgactt tcctggacga gcggagaagt cgatatgcgg 1200
 ttacaggaata tcatgcgtgc cattcatcat tcctgcaccg aatacggcaa gaagcctgac 1260
 ggtacgggtca actatgtgga tgggtgccaat gttgccggat ttgtgaaagt ggccgaggca 1320
 atgctggcgc aaggggtgat ctgataa 1347

<210> 16

<211> 447

<212> PRT

<213> Nitrosomonas europaea

<400> 16

Met Lys Tyr Asn Ser Ile Glu Glu Phe Lys Asn Tyr Val Ser Glu Arg
 1 5 10 15
 Asn Pro Gly Gln Pro Glu Phe Leu Gln Ala Val Ser Glu Val Ile Glu
 20 25 30
 Ser Leu Trp Pro Phe Ile Val Asp His Ser Arg Tyr Ala Glu Gln Gly
 35 40 45
 Leu Leu Asp Arg Leu Ile Glu Pro Glu Arg Met Ile Ile Phe Arg Val
 50 55 60
 Ala Trp Val Asp Asp Arg Gly Glu Val Lys Val Asn Arg Gly Tyr Arg
 65 70 75 80
 Ile Gln Tyr Asn Ser Ala Ile Gly Pro Tyr Lys Gly Gly Thr Arg Phe
 85 90 95
 His Pro Ser Val Asn Leu Ser Ile Leu Lys Phe Leu Ala Phe Glu Gln
 100 105 110
 Thr Phe Lys Asn Ala Leu Thr Thr Leu Pro Met Gly Gly Gly Lys Gly
 115 120 125
 Gly Ser Asp Phe Asp Pro Lys Gly Lys Ser Pro Gly Glu Ile Met Arg
 130 135 140
 Phe Cys Gln Ala Tyr Ala Ala Glu Leu Phe Arg His Val Gly Ala Asp
 145 150 155 160
 Thr Asp Val Pro Ala Gly Asp Ile Gly Val Gly Gly Arg Glu Val Gly
 165 170 175
 Tyr Met Ala Gly Met Val Lys Lys Leu Thr Asn Arg Ser Asp Cys Val

